

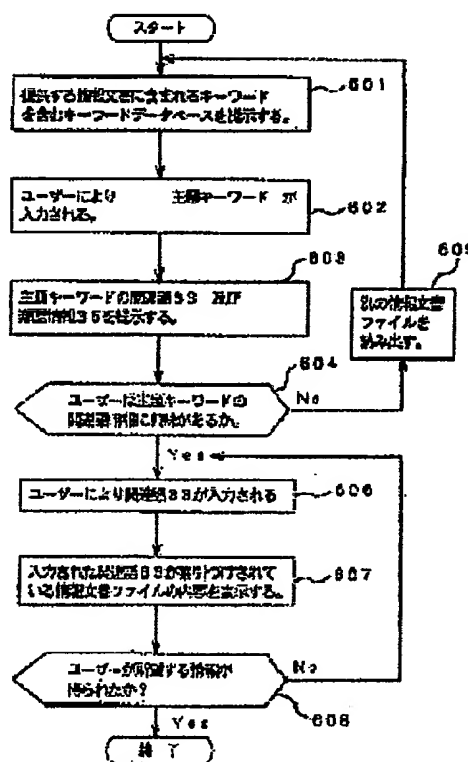
INFORMATION GENERATION SYSTEM, INFORMATION RETRIEVAL SYSTEM AND RECORDING MEDIUM

Patent number: JP2000259671
Publication date: 2000-09-22
Inventor: FUJIOKA TAKAKO
Applicant: DAINIPPON PRINTING CO LTD
Classification:
 - international: G06F17/30
 - european:
Application number: JP19990067318 19990312
Priority number(s): JP19990067318 19990312

Report a data error here

Abstract of JP2000259671

PROBLEM TO BE SOLVED: To provide an information generation system capable of indexing a document so as to easily identify information included in the document and to provide an information retrieval system to which a user's viewpoint in retrieving work can be inputted and which has high efficiency of retrieving work. **SOLUTION:** A computer presents a keyword database including keywords included in an information document to be provided (step 601). A user selects and inputs a keyword in which the user is interested (step 602). The computer presents related work and synonym information corresponding to the inputted keyword (step 603). The user refers to the presented related work information and inputs a keyword of the user's interest (steps 605, 606). The computer displays the contents of a document file to which the inputted related word is indexed on its display (step 607).



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2000-259671

(P2000-259671A)

(43) 公開日 平成12年9月22日 (2000.9.22)

(51) Int.Cl.

識別記号

G 0 6 F 17/30

F I

特開ト* (参考)

C 0 6 F 15/401

3 1 0 A 5 B 0 7 U

15/40

3 7 0 A

15/403

3 2 0 D

3 4 0 B

審査請求 未請求 請求項の数 7 O L (全 10 頁)

(21) 出願番号

特願平11-67318

(22) 出願日

平成11年3月12日 (1999.3.12)

(71) 出願人 000002897

大日本印刷株式会社

東京都新宿区市谷加賀町一丁目1番1号

(72) 発明者 藤岡 孝子

東京都新宿区市谷加賀町一丁目1番1号

大日本印刷株式会社内

(74) 代理人 100096091

弁理士 井上 誠一

Fターム (参考) 5B075 ND03 NK32 NK35 NR02 NR03

NR20 PP22 PP28 PP30 PR10

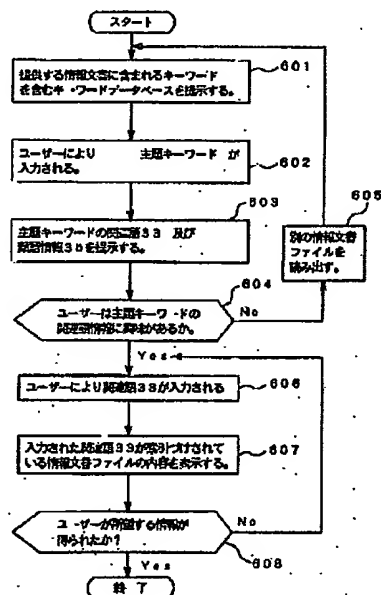
QM05 QP03

(54) 【発明の名称】 情報生成システム、情報検索システム、及び記録媒体

(57) 【要約】

【課題】 文書内の情報を容易に判別できるように文書に索引付けする情報生成システム、及び検索作業においてユーザの視点が入力でき、検索作業効率の高い情報検索システムを提供すること。

【解決手段】 コンピュータ5は提供する情報文書に含まれるキーワードを含むキーワードデータベース11を提示する (ステップ601)。ユーザ10は興味のあるキーワードを選択し、入力する (ステップ602)。コンピュータ5は、入力されたキーワードの関連語33及び類語情報35を提示する (ステップ603)。ユーザ10は提示されたこれらの関連語情報を参照し、興味のある関連語33を入力する (ステップ605、ステップ606)。コンピュータ5は、入力された関連語33が索引付けされている文書ファイルの内容をコンピュータ9のディスプレイ上に表示する (ステップ607)。



【特許請求の範囲】

【請求項1】 文書からキーワードを抽出する手段と、前記文書において抽出されたキーワードの近傍から、前記キーワードに対応する関連語及び関連語の類似語を抽出する手段と、抽出された関連語及び関連語の類似語の中から代表語を決定する決定手段と、前記決定された代表語を前記文書の索引として付加する索引付加手段と、を具備することを特徴とする情報生成システム。

【請求項2】 前記決定手段は、前記文書のキーワード近傍に出現する頻度の最も高い関連語又は関連語の類似語を代表語として決定することを特徴とする請求項1記載の情報生成システム。

【請求項3】 情報を検索するための情報検索システムであって、キーワードに対応する関連語情報を保持する関連語情報保持手段と、前記関連語情報が索引として付加された文書を保持する文書保持手段と、ユーザがキーワードを入力すると、前記関連語情報保持手段からこのキーワードに関連する関連語情報を抽出する手段と、抽出された関連語情報のうちユーザが選択した関連語情報が索引として付加された文書を前記文書保持手段から検索する手段と、を具備することを特徴とする情報検索システム。

【請求項4】 情報を検索するための情報検索システムであって、キーワードを保持するキーワード保持手段と、前記キーワードに対応する関連語情報を保持する関連語情報保持手段と、前記関連語情報が索引として付加された文書を保持する文書保持手段と、前記キーワード保持手段に保持されたキーワードを表示する手段と、ユーザが表示されたキーワードから所定のキーワードを選択し、入力すると、前記関連語情報保持手段からこのキーワードに関連する関連語情報を抽出する手段と、抽出された関連語情報のうちユーザが選択した関連語情報が索引として付加された文書を前記文書保持手段から検索する手段と、を具備することを特徴とする情報検索システム。

【請求項5】 コンピュータを、文書からキーワードを抽出する手段と、前記文書において抽出されたキーワードの近傍から、前記キーワードに対応する関連語及び関連語の類似語を抽出する手段と、抽出された関連語及び関連語の類似語の中から代表語を決定する決定手段と、

前記決定された代表語を前記文書の索引として付加する索引付加手段と、

して機能させるためのプログラムを記録する記録媒体。

【請求項6】 コンピュータを、キーワードに対応する関連語情報を保持する関連語情報保持手段と、前記関連語情報が索引として付加された文書を保持する文書保持手段と、ユーザがキーワードを入力すると、前記関連語情報保持手段からこのキーワードに関連する関連語情報を抽出する手段と、

抽出された関連語情報のうちユーザが選択した関連語情報が索引として付加された文書を前記文書保持手段から検索する手段と、

して機能させるためのプログラムを記録する記録媒体。

【請求項7】 コンピュータを、キーワードを保持するキーワード保持手段と、前記キーワードに対応する関連語情報を保持する関連語情報保持手段と、前記関連語情報が索引として付加された文書を保持する文書保持手段と、前記キーワード保持手段に保持されたキーワードを表示する手段と、

ユーザが表示されたキーワードから所定のキーワードを選択し、入力すると、前記関連語情報保持手段からこのキーワードに関連する関連語情報を抽出する手段と、抽出された関連語情報のうちユーザが選択した関連語情報が索引として付加された文書を前記文書保持手段から検索する手段と、

して機能させるためのプログラムを記録する記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、情報生成システム、情報検索システム及び記録媒体に関するものである。

【0002】

【従来の技術】昨今、インターネットや通信ネットワーク等を通じて、文書情報データベースから情報を検索することが多い。通常の情報検索では、ユーザがキーワード等の条件を入力することで関連情報が出力され、ユーザによる条件の追加や条件のAND/ORを採ることにより、情報が絞り込まれ、ユーザの所望の情報が得られる。

【0003】

【発明が解決しようとする課題】しかしながら、このようなキーワードによる情報検索では、ユーザがどのような視点でそのキーワードを選択したかが考慮されず、入力したキーワードに関する膨大な情報が出力され、ユーザが所望する情報を得るまでに手間がかかるという問題がある。

【0004】本発明はこのような問題に鑑みてなされたもので、その目的とするところは、文書内の情報を容易に判別できるように文書に索引付けする情報生成システム、及び検索作業においてユーザの視点が入力でき、検索作業効率の高い情報検索システムを提供することにある。

【0005】

【課題を解決するための手段】前述した目的を達成するために第1の発明は、文書からキーワードを抽出する手段と、前記文書において抽出されたキーワードの近傍から、前記キーワードに対応する関連語及び関連語の類似語を抽出する手段と、抽出された関連語及び関連語の類似語の中から代表語を決定する決定手段と、前記決定された代表語を前記文書の索引として付加する索引付加手段とを具備することを特徴とする情報生成システムである。

【0006】第2の発明は、情報を検索するための情報検索システムであって、キーワードに対応する関連語情報を保持する関連語情報保持手段と、前記関連語情報が索引として付加された文書を保持する文書保持手段と、ユーザがキーワードを入力すると、前記関連語情報保持手段からこのキーワードに関連する関連語情報を抽出する手段と、抽出された関連語情報のうちユーザが選択した関連語情報が索引として付加された文書を前記文書保持手段から検索する手段とを具備することを特徴とする情報検索システムである。

【0007】第3の発明は、情報を検索するための情報検索システムであって、キーワードを保持するキーワード保持手段と、前記キーワードに対応する関連語情報を保持する関連語情報保持手段と、前記関連語情報が索引として付加された文書を保持する文書保持手段と、前記キーワード保持手段に保持されたキーワードを表示する手段と、ユーザが表示されたキーワードから所定のキーワードを選択し、入力すると、前記関連語情報保持手段からこのキーワードに関連する関連語情報を抽出する手段と、抽出された関連語情報のうちユーザが選択した関連語情報が索引として付加された文書を前記文書保持手段から検索する手段とを具備することを特徴とする情報検索システムである。

【0008】

【発明の実施の形態】以下、図面に基づいて本発明の実施の形態を詳細に説明する。図1は、本実施の形態に係るシステムの概略構成を示す図である。

【0009】図1において、コンピュータ5にはキーワードデータベース11、関連語情報データベース13、及び文書データベース15-1が設けられる。コンピュータ7はネットワーク3を介してコンピュータ5に接続され、コンピュータ7には文書データベース15-2が設けられる。

【0010】キーワードデータベース11は概念的に階

層構造をとるキーワード群を保持する。関連語情報データベース13は検索時の視点の情報となる関連語情報を保持する。文書データベース15-1、15-2は、解析未処理の文書または解析処理済みの文書を保持する。

【0011】コンピュータ9はユーザ10が検索に使用するためのコンピュータである。コンピュータ9にはユーザ10により情報検索のために必要なキーワード等の入力される。

【0012】コンピュータ5は入力されたキーワード等に適合する文書を文書データベース15-1から読みだし、コンピュータ9に出力する。また、コンピュータ5は、ネットワーク3を介して、コンピュータ7に管理される文書データベース15-2に保持される文書の検索も行う。

【0013】CD-ROM17はコンピュータ5に後述するような索引付け処理、及び情報検索処理を行わせるためのプログラムが記録される。また、CD-ROM17には索引付けを行った文書データも記録される。尚、CD-ROM17に記録されたプログラム及び文書データは磁気ディスク等の他の記録媒体に格納されてもよい。

【0014】図2はキーワードデータベース11に保持されるキーワードの構成を示す図である。キーワードデータベース11には、あらゆる分野別に名詞がその概念ごとに階層型に登録されている。以下の説明は「食材」を例にとって行う。

【0015】図2に示すように、上位概念である「食材」の内、上位語23は例えば「乳製品」、「肉類」、…といったものがある。その上位語23の下には、概念的にその上位語23に属するキーワード25が登録される。例えば、上位語「乳製品」には、「チーズ」、「バター」といったキーワード25が登録される。

【0016】図3は関連語情報データベース13に保持される関連語情報の一例を示す図である。関連語情報とは図3に示す関連語33とそれに属する類語情報35である。関連語33は、ユーザ10がキーワード25に関する情報を検索する場合に、どのような情報を所望しているか、即ち視点を入力するために設けられるものである。

【0017】従って、ユーザ10のキーワード25に対する視点が明確になるように、関連語33には動詞が選ばれる。例えば、キーワード25が「チーズ」である場合、関連語33はユーザ10が「チーズ」をどうしたいか、即ち「チーズ」に対する消費行動を示すものであり、図3に示すように「使う」、「買う」、「食べる」、「知る」等がある。

【0018】類語情報35は、ユーザ10が自分の所望する情報に適した関連語33を選択するために参照とする情報であり、関連語33に属する。例えば、関連語「使う」に対する類語情報35は「料理、調理、切り

方、保存法製品、製法、原材料、原産国、地方、輸入、…」等がある。このように類語情報35を参照すれば、ユーザ10は関連語「使う」の意味を明確に判断でき、適切な関連語33を選択することができる。

【0019】尚、関連語情報データベース13において、関連語33及び類語情報35はキーワード25に対応させて登録される。

【0020】次に、情報生成システムによる文書の索引付けについて説明する。図4は文書に関連語33を索引として付加する作業の手順を示すフローチャートである。コンピュータ5は、文書データベース15-1等に保持された文書ファイルから情報文書ファイルを読み出して、この情報文書ファイルをテキストデータとして入力する(ステップ401)。

【0021】コンピュータ5は、文書中のテキストを形態素解析して名詞を抽出し、この名詞の内キーワードとしてキーワードデータベース11に登録されている名詞に対して以下の処理を行う(ステップ402)。

【0022】抽出された名詞、即ちキーワードに関して、キーワードデータベース11に当該キーワードの関連語情報が存在するかどうかを判定する(ステップ403)。関連語情報がない場合、文書中のキーワード近傍において出現頻度の高い単語を関連語情報として登録する(ステップ404)。

【0023】抽出された名詞、即ちキーワードに関連語情報が存在する場合、文書中のキーワードの近傍に関連語又は類似語が出現しているかを調べる(ステップ405)。関連語と類似語のグループ毎に出現頻度を調べ、頻度の高いグループを選ぶ(ステップ406)。

【0024】次に、選択されたグループの代表語、即ち関連語33を索引としてテキスト中のキーワード位置に付加する(ステップ407)。ステップ402で抽出した名詞の内、キーワードとして登録されているものは残っていないかを判断し(ステップ408)、残っているキーワードに対して同様にステップ403からステップ408の作業を繰り返し、索引付けを行う。

【0025】以下、図4に示す索引付けの手順を詳細に説明する。図5は索引を付ける文書51と関連語33を示す図である。図1に示す文書データベース15-1から取り出された文書51は、そのテキストの形態素解析が行われ、名詞のみが抽出される。例えば、「ブルーチーズソースの作り方」というテキストから「ブルーチーズ」、「ソース」、「作り方」が取り出される。

【0026】抽出された名詞がキーワードデータベース11に登録されている場合、そのキーワードに属する関連語情報を取り出す。例えば、「ブルーチーズ」はキーワード25-1の「チーズ」としてキーワードデータベース11に登録されており、その関連語情報として図3に示す関連語33と類語情報35がある。

【0027】図5に示す文書51において、キーワード

「チーズ」の近傍501、502、503において出現する「作り方」、「材料」、「手順」は、図3に示す関連語「使う」に属する類語情報35に含まれる単語そのもの、または類語情報に含まれる単語の類似語である。

【0028】従って、文書51には関連語「使う」及びその類似語のグループに属する単語の出現頻度が高く、選択されたグループの代表語である「使う」が文書51の索引としてテキスト中のキーワード25-1の位置に付加される。こうして索引が付けられた文書51は文書データベース15-1等に格納される。

【0029】このように本実施の形態によれば、文書中に出現するキーワードを抽出し、この文書がキーワードに対するどのような情報に関するものであるかを表すための索引を付けることができる。

【0030】次に、情報検索システム1による検索処理について説明する。情報検索システム1は、上記のように索引付けられた文書からユーザ10の所望する情報を検索するものである。図6は情報検索処理の手順を示すフローチャートである。

【0031】コンピュータ5は提供する情報文書に含まれるキーワードを含むキーワードデータベース11を提示する(ステップ601)。例えば、図2に示すような樹形図等を表示させる。これは、コンピュータ5が文書データベース15-1に保持し、ユーザ10に提供可能な情報文書に含まれるキーワードとキーワードデータベース11におけるその周辺のキーワード群を、情報文書ファイルの目次情報として提示することである。

【0032】ここで、コンピュータ5から提示されるキーワードとキーワードデータベース11におけるその周辺のキーワード群はネットワーク3を介してユーザ入力用コンピュータ9のディスプレイ上に表示される。以下、コンピュータ5が表示した内容は全てコンピュータ9のディスプレイ上に表示されるものとする。

【0033】ユーザ10は提示されたキーワード群から興味のあるキーワードを選択し、入力する(ステップ602)。ここで、ユーザ10によって入力されたキーワードを主題キーワードとする。尚、ステップ601においてキーワードが提示されない場合、ユーザ10がキーワードを直接入力することもできる。

【0034】コンピュータ5は、コンピュータ9のディスプレイ等に主題キーワードの関連語33及び類語情報35を提示する(ステップ603)。ユーザ10は提示されたこれらの関連語情報を参照し、これらに興味があるかどうかを判断する(ステップ605)。

【0035】ユーザ10が関連語情報に興味がない場合は、別の情報文書ファイルを読みだし(ステップ605)、これまでの処理を繰り返す。ユーザ10が関連語情報に興味のある場合、ユーザ10は類語情報35を参照して興味のある関連語33を入力する(ステップ606)。

【0036】次に、コンピュータ5は、入力された関連語33が索引付けされている文書ファイルの内容をコンピュータ9のディスプレイ上に表示する（ステップ607）。ユーザ10はその内容を確認し、ユーザ10の所望する情報が得られたかどうかを判定する（ステップ608）。ユーザ10が所望する情報を得られない場合、ユーザ10は再度関連語33を入力する（ステップ606）。

【0037】ユーザ10は、ステップ606からステップ608の作業を繰り返し、情報検索の視点となる関連語33を入力することで、所望の文書情報を得ることができる。

【0038】図7及び図8は、関連語33が索引付けされた情報文書ファイルを示す図である。例えば、主題キーワード61が「チーズ」と入力されると（ステップ602）、「チーズ」の関連語33である「使う」、「食べる」、「買う」等が表示される（ステップ603）。

【0039】文書データベース15-1等に格納される原文書ファイル71-1、71-2、71-3には関連語33が索引として付けられている。例えば、ユーザ10が「チーズ」を「食べる」に関する情報が知りたい場合、ユーザ10が関連語33として「食べる」を入力すると（ステップ606）、「食べる」が索引付けされている原文書ファイル71-2の内容がコンピュータ9にディスプレイ上に表示される（ステップ607）。

【0040】また、情報文書ファイルには図8に示すような情報ファイル等がある。例えば、関連語「使う」が索引として付けられているものには、例えば「チーズ」に関する出版物の情報を示す情報ファイル81-1がある。

【0041】また、「知る」が索引付けされている情報ファイルには店舗（学校、教室）をまとめた情報ファイル81-2があり、「食べる」が索引付けされている情報ファイルには例えば美味しい「チーズ」の食べられる店舗（レストラン）に関する情報をまとめた情報ファイル81-3がある。

【0042】また、「買う」が索引付けされている情報ファイルには、「チーズ」を取り扱う通販業者（食材）情報ファイル81-4や店舗（食材店）情報ファイル81-5等がある。

【0043】このように本実施の形態によれば、ユーザはキーワードとそのキーワードに対する関連語33を入

力することで情報検索の視点や目的を入力でき、目的にあった情報を容易に得ることができる。

【0044】尚、このような検索作業で、ユーザによって入力された主題キーワード61や関連語33は蓄積され、マーケティング情報等として活用することもできる。また、情報文書ファイルはコンピュータ5とネットワーク3で接続されたコンピュータ7が有する文書データベース15-2に格納されていてもよい。

【0045】また、CD-ROM等の記録媒体に、キーワードデータベース11や関連語情報データベース13の内容及び情報文書ファイルを格納し、ユーザに提供することも可能である。

【0046】尚、本出願人は、関連語情報データベース13の構築に関して別途特許出願を行っている。

【0047】

【発明の効果】以上、詳細に説明したように本発明によれば、情報内容にしたがって文書を容易に分類でき、情報検索作業における作業効率を向上することができる。

【図面の簡単な説明】

【図1】 本発明の1実施の形態に係るシステムを示す図

【図2】 キーワードデータベース11を示す図

【図3】 関連語情報データベース13を示す図

【図4】 文書への索引付け作業を示すフローチャート

【図5】 文書51と関連語33を示す図

【図6】 情報検索作業の手順を示すフローチャート

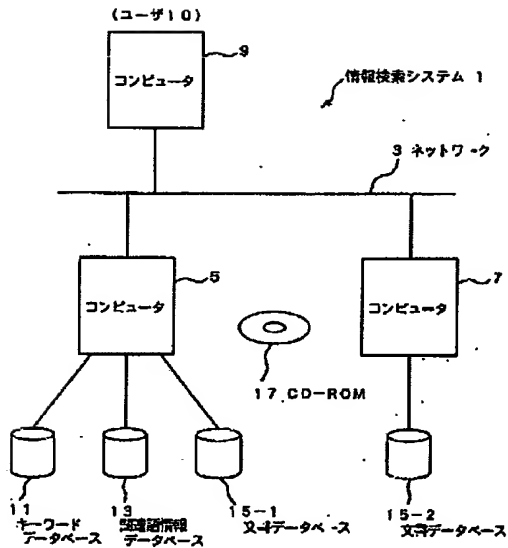
【図7】 関連語が索引付けされた原文書ファイル71を示す図

【図8】 関連語が索引付けされた情報ファイル81を示す図

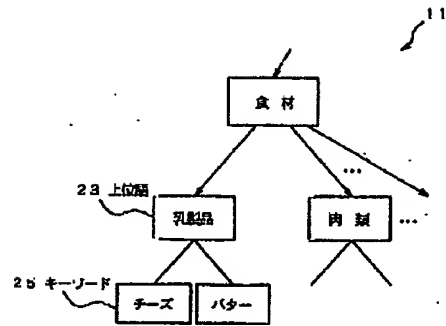
【符号の説明】

- 1.....情報検索システム
- 3.....ネットワーク
- 5.....コンピュータ
- 7.....コンピュータ
- 9.....コンピュータ
- 10.....ユーザ
- 11.....キーワードデータベース
- 13.....関連語情報データベース
- 15-1、15-2.....文書データベース
- 17.....CD-ROM

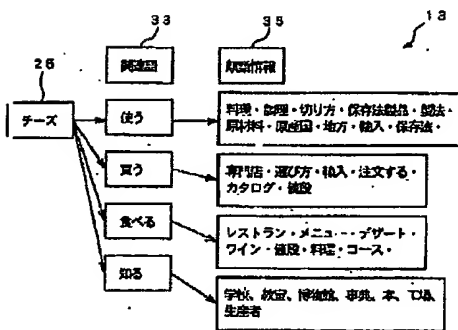
【図1】



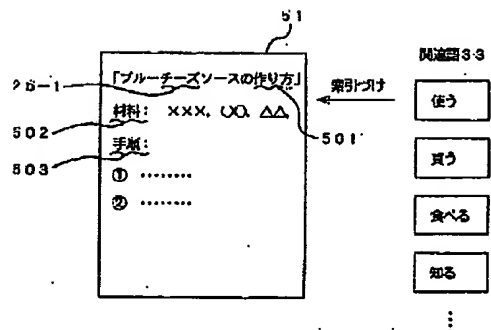
【図2】



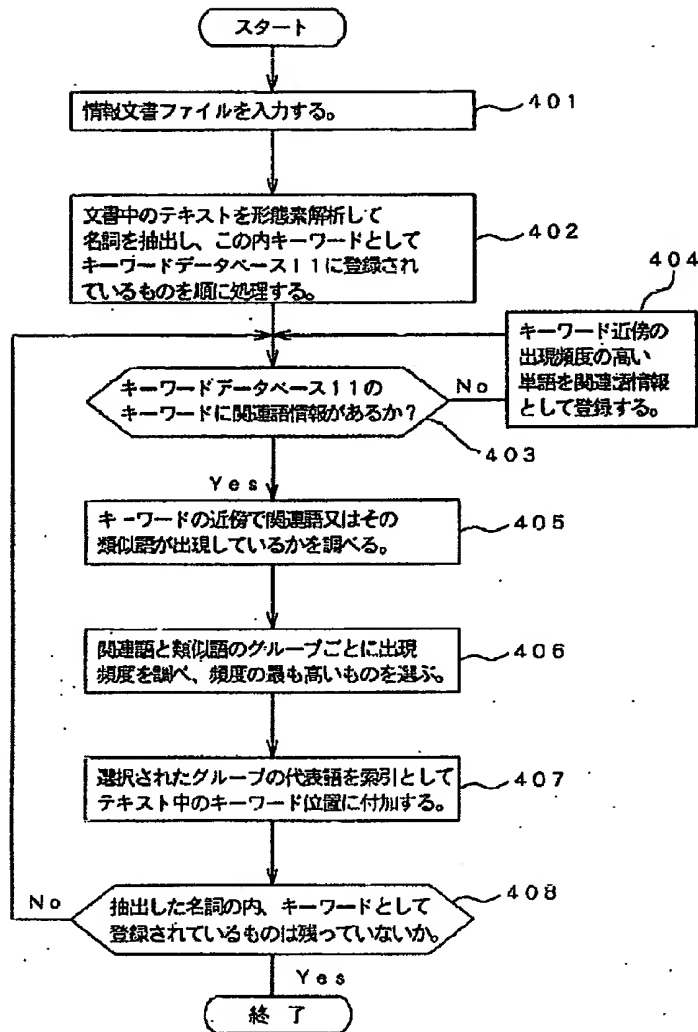
【図3】



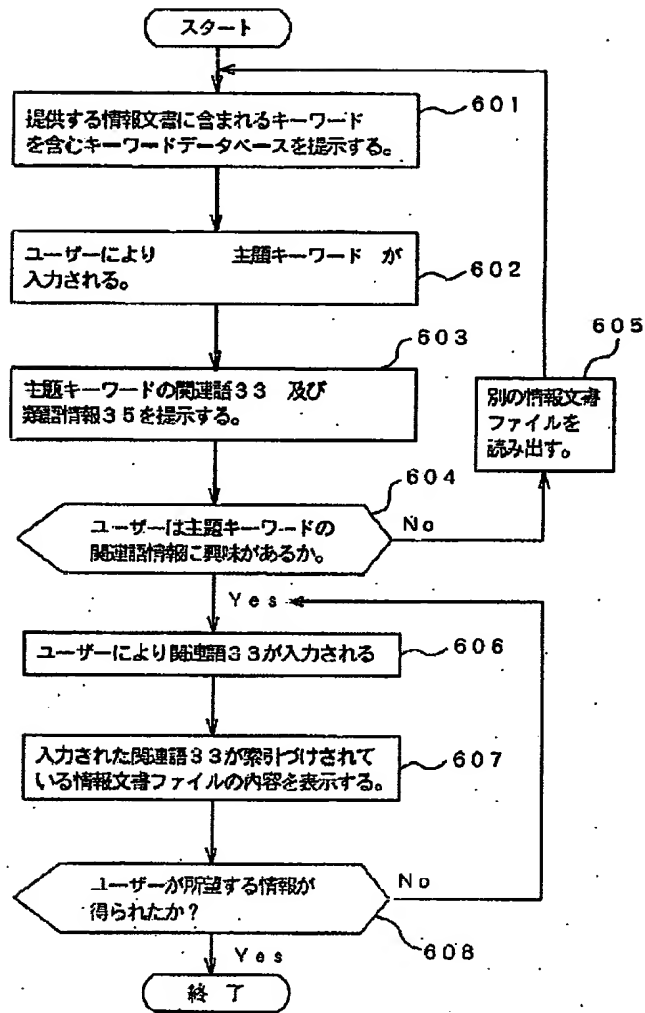
【図5】



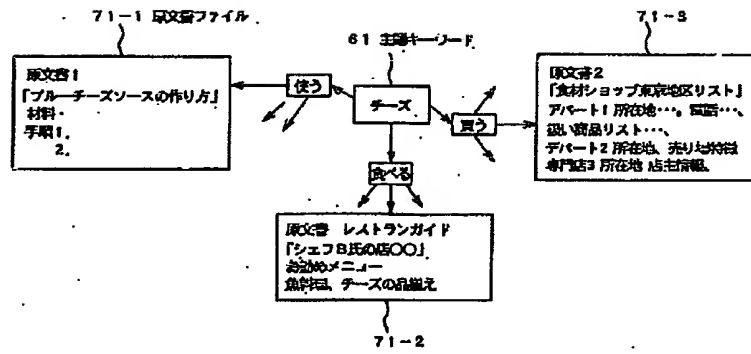
【図4】



【図6】



【図7】



【図8】

